DOCUMENT RESUME

ED 190 076 IR 008 557

AUTHOR Hoover, Ryan E.
TITLE Database Overlap vs. Complementary Coverage in Forestry and Forest Products: Factors in Database Acquisition.
PUB DATE Mar 80
NOTE 26p.: Paper presented at the National Online Information Meeting (New York, NY, March 25-27, 1980).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Citations (References): *Data Bases: *Data Collection: *Forestry: *Information Services: Information Systems: Lumber Industry: *Online Systems

ABSTRACT

This study examines (1) subject content, (2) file size, (3) types of documents indexed, (4) range of years spanned, and (5) level of indexing and abstracting in five databases which collectively provide extensive coverage of the forestry and forest products industries: AGRICOLA, CAB ABSTRACTS, FOREST PRODUCTS (AIDS), PAPERCHEM, and PIRA. The extent to which these databases overlap and complement each other and the ways in which their characteristics affect database acquisition by a commercial online vendor are explored. (FM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

SP-4053

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED190076

DATABASE OVERLAP VS. COMPLEMENTARY COVERAGE IN FORESTRY AND FOREST PRODUCTS: FACTORS IN DATABASE ACQUISITION

By

Ryan E. Hoover

Food and Agricultural Information Specialist

SDC® Search Service

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Ryan E. Hoover

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Presented at

National Online Meeting

New York

March 1980

IR008357

428

ABSTRACT

Bibliographic databases exhibit both overlapping and complementary coverage. There are five databases which, collectively, provide extensive coverage of the forestry and forest products industries. These databases are 1) AGRICOLA, 2) CAB ABSTRACTS, 3) FOREST PRODUCTS (AIDS), 4) PAPERCHEM, and 5) PIRA. The forestry and forest products industries are broken into three major segments: 1) Forestry, 2) Wood/Forest Products, and 3) Pulp/Paper Products. The five databases show overlapping and complementary coverage across these industrial segments in 1) subject content; 2) file size (number of documents indexed); 3) types of documents indexed; 4) range of years spanned; and 5) the level of indexing and abstracting. These factors are used in an attempt to show the extent to which databases overlap and complement each other and to indicate how these factors affect database acquisition by a commercial online vendor. Some conclusions regarding database overlap and bibliographic control are made.

INTRODUCTION

Bibliographic databases which are available in machine-readable format exhibit both overlapping and complementary coverage of various topics. Traditionally, in the days when bibliographic databases existed only in printed form as index and abstract journals, moderate overlap with other such journals was not considered to be much of a problem. Typically, a searcher used only one printed index/abstract journal to compile a bibliography. The usual approach was to peruse the source which provided the most extensive coverage of the topic of interest.

In the early days of online bibliographic searching, overlap was not a serious problem, since only a few databases existed online. Today, approximately 150 databases are readily available on online search systems and a typical bibliographic searcher seldom limits a search to just one database. Cross-searching of multiple databases is easy on online systems and, while this insures retrieval of the widest possible coverage of a topic, considerable duplication of retrieved documents sometimes occurs. A few database producers, e.g., Biological Abstracts, Chemical Abstracts, and Engineering Index, have made attempts in recent years to more carefully define their abstracting

areas in an attempt to reduce overlap. Most database suppliers, however, continue to index documents in their traditional areas, often overlapping in the coverage of the same literature in other databases.

A vendor of online bibliographic search services analyzes overlapping and complementary coverage before adding new databases to the existing offerings. An attempt is made to determine the degree of overlap with other existing databases, and also to determine how well a potential database complements the others by partially or completely filling a gap in coverage not provided by databases already on the system. In addition to examining the database-specific factors, e.g., subject content, file size, years covered, etc., a vendor will consider other factors, such as the needs of target customer groups, the economic performance of related databases on the system, etc. This process of analysis is discussed for five databases covering the literature in the broad subject area of forestry and forest products.

DATABASES IN FORESTRY AND FOREST PRODUCTS

There are five bibliographic databases which cover the literature in the areas of forestry and the forest products industries. These are:

1) AGRICOLA, produced by the U.S. Dept. of Agriculture's SEA/TIS;

2) CAB ABSTRACTS, produced by the Commonwealth Agricultural Bureaux (UK);

3) FOREST PRODUCTS (AIDS), produced by the Forest Products Research Society (USA);

4) PAPERCHEM, produced by the Institute of Paper Chemistry (USA);

5) PIRA, produced by the Research Association for the Paper and Board, Printing and Packaging Industries (UK).

Collectively, these five databases provide extensive coverage of literature in forestry and the forest products industries, consisting of all wood and paper products manufacturing. The forestry and forest product industries may be segmented into three broad areas:

1) Forestry -- the science and industry of natural and man-made forests including silviculture, planting, mensuration, harvesting, transporting, etc., of raw timber and wood;

2) Wood and Forest Products -- the science and technology of lumber and wood processing, cutting, shaping, grinding,

gluing, construction, etc., of any products made from wood, exclusive of pulp and pulping;

3) Pulp and Paper Products -- the science and technology of wood pulping, paper manufacture, and manufacture of related products from wood and cellulose pulp.

OVERLAPPING AND COMPLEMENTARY COVERAGE

The AGRICOLA, CAB ABSTRACTS, FOREST PRODUCTS, PAPERCHEM and PIRA databases both overlap and complement each other on the topics of forestry, wood/forest products, and pulp/paper products. This is exhibited in the following database-specific areas:

1) Subject content;

2) File size or number of documents indexed;

3) The types of documents indexed;

4) The range of years spanned;

5) The level of indexing and abstracting.

A rough analysis of each of these factors is presented from the point of view of an online vendor deciding whether or not to provide these databases online.

SUBJECT CONTENT

While all five databases provide at least some coverage of forestry, wood/forest products, and pulp/paper products, each has a primary emphasis on only one of the three areas. AGRICOLA provides extensive primary coverage of forestry and secondary emphasis on forest industries and products. Appendix I lists the coverage of these areas by primary category codes. Emphasis is on documents from the United States and Canada, although world-wide literature is covered. CAB ABSTRACTS is more international in scope than AGRICOLA and it, too, primarily covers forestry with secondary coverage of forest products. Appendix II lists CAB ABSTRACTS coverage by file segment (abstract journal) and by Oxford Decimal Class (Oxford System of Decimal Classification for Forestry). CAB ABSTRACTS consists of 28 individual abstract journals, searchable as file segments or subfiles. The two covering forestry are Forestry Abstracts (FA - journal Øf) and Forest Products Abstracts (FPA - journal 1F). Both AGRICOLA and CAB ABSTRACTS provide minimal coverage of the pulp and paper industries.

FOREST PRODUCTS (AIDS) and PAPERCHEM are devoted almost exclusively to wood/forest products and pulp/paper products, respectively, and are highly complementary. Both cover forestry to

a very small extent and both overlap into each other's primary coverage areas only marginally. FOREST PRODUCTS both overlaps and complements AGRICOLA and CAB ABSTRACTS but is particularly thorough in its coverage of wood milling and machining. It is worldwide in scope but only about 10% of original documents are in a non-English language. Appendix III lists the FOREST PRODUCTS subject content by category code. PAPERCHEM provides very comprehensive worldwide coverage of the pulp/paper industries with emphasis on the chemistry of pulp and paper manufacturing. Appendix IV lists the subject content of PAPERCHEM. PIRA consists of four individual abstract journals, of which only one is of interest for this paper: Paper and Board Abstracts (PBA). Because of the large size of PAPERCHEM (ca. 134,000 records) and the small size of PIRA/PBA (ca. 12,000 records), it is doubtful that there is much in PIRA not covered by PAPERCHEM. PIRA contains little on the chemistry of pulp, paper, and board manufacturing, although other subfiles include considerable material on the printing and packaging industries, a logical extension of paper products.

FILE SIZE

The number of documents related to forestry or forest products indexed by each database is an important factor. File size,

alone, does not indicate the quality of worth of a database, however. An attempt was made to determine the approximate number of documents in each of the five databases which are indexed to forestry, wood/forest products, and pulp/paper products. Appendixes VI-X show how the citation counts were derived for each database. It should be noted that the breakdowns are very general and serve only to indicate the relative size and subject overlap of the five databases. Appendix XI lists the citation counts by industry segment and Appendix XII shows the relative sizes visually. These results do not indicate types of documents covered, the span of years covered, the level of indexing or the presence of abstracts, etc.

AGRICOLA is the largest file covering forestry and forest industries, while CAB ABSTRACTS is next in size. FOREST PRODUCTS is a small file, comparatively, which provides coverage of wood/forest products in content (wood milling and machining), in types of documents, and in years spanned not found in either AGRICOLA or CAB ABSTRACTS. Obviously, PAPERCHEM is the file to search for pulp/paper products in terms of size. PIRA is also good in pulp/paper (exclusive of chemistry) but it almost totally overlaps PAPERCHEM's coverage of literature in this area.

TYPES OF DOCUMENTS INDEXED

Each of the five databases indexes one or two document types not indexed by the others. Appendix XIII shows the document types covered by each database, as far as could be determined from database documentation. Results indicate that certain databases should be searched for certain document types. For example, U.S. government documents are indexes by AGRICOLA and FOREST PRODUCTS and theses and dissertations are indexed by CAB ABSTRACTS and PAPERCHEM. All five databases cover the major documents types: journals, books, conference proceedings and research or technical reports.

RANGE OF YEARS COVERED

Each of the five databases being considered differs in the span of years covered. AGRICOLA and CAB ABSTRACTS provide retrospective documentation back to 1970 and 1972, respectively. FOREST ABSTRACTS provides limited coverage (of all documents published by the Forest Products Research Society) back to 1945. PAPERCHEM covers the literature back to 1968 while PIRA documents literature only from 1975. Appendix XIV graphically shows the years covered by the five databases. There is considerable chronological overlap between AGRICOLA and CAB ABSTRACTS in forestry and wood/forest products while FOREST PRODUCTS covers

early years not documented by the two former databases. There is complete overlap, chronologically, of PIRA by PAPERCHEM.

LEVEL OF INDEXING AND ABSTRACTING

The quality and depth of indexing varies from database to database, and, sometimes within a database if it contains indexing and abstracting from different sources. Documents in AGRICOLA exhibit differing levels of indexing. Most are fairly adequately indexed using free terms from titles and the subject headings index of the Bibliography of Agriculture, (published as Agricultural Terms, 2nd edition, Phoenix, Oryx Press (1978)). Forestry and forest industries entries in AGRICOLA are not abstracted (abstracts occur only in the Food and Nutrition, (FNC) and Agricultural Economics (AGE) subfiles).

Records in CAB ABSTRACTS (Forestry Abstracts (FA) and Forest Products Abstracts (FPA) are indexed according to Terminology of Forest Science, Technology, and Practice, and Products, (the multilingual forestry terminology series, no.1), Washington, D.C., Society of American Foresters. All forestry and forest products entries are also classified using the Oxford Decimal Classification (ODC) for forestry (see Appendix II). Most entries include thorough abstracts, as well.

FOREST PRODUCTS is an online version of the FPRS Abstract Information Digest Service (AIDS), published in microfiche. Documents are well indexed using an in-house thesaurus and a detailed set of subject category codes (Appendix III), but do not include abstracts online. Abstracts are available on the microfiche AIDS product, however, and FPRS indicates that they may include the abstracts on tape beginning in 1980. PAPERCHEM is very extensively indexed and abstracted and uses the Thesaurus of Pulp and Paper Terms, Point Claire, Quebec, Pulp and Paper Research Institute of Canada, 1978. PIRA is indexed using chapter subject headings in Paper and Board Abstracts; these descriptors tend to be rather general, although PIRA records include abstracts.

CONCLUSIONS

This somewhat cursory analysis would indicate that at least four of the databases discussed--AGRICOLA, CAB ABSTRACTS, FOREST PRODUCTS, and PAPERCHEM--are important for the best possible coverage of literature in forestry, wood/forest products, and pulp/paper products. PIRA is important if packaging and printing are also of interest. In addition to the foregoing, an online vendor also considers some additional factors before deciding to add new databases to its online offering:

1) the needs of its target customer or user group;

2) the economic performance of existing related files;

3) the overall quality of the data.

A survey of customers in the wood, forest products, and paper industries indicate that it would be desirable to have all of the databases discussed. If existing files are doing well, i.e., there is sufficient traffic so that the databases at least pay for themselves, then it is probably worthwhile to add additional related files. The entire family of databases would probably stimulate traffic on each other, since they would collectively, provide "one-vendor" searching in the areas of forestry and forest products. Finally, after looking at a test tape of data from a new database supplier, a decision to add or not to add a file may be made depending on the overall quality of data. If it is "sloppy", with little evidence of quality control, it may not be desirable to add it to the system. Among the databases discussed, all appear to be quality files, each containing sufficient unique material to be included in an online system.

The need exists for all database suppliers to cooperate with each other increasingly as more databases appear in machine-readable form. Chemical Abstracts, Biological Abstracts, and

Engineering Index have pioneered in defining their areas of coverage to avoid extensive overlap. As this is done increasingly in the coming years, some small databases may be absorbed by others, mergers may occur, and some databases might even disappear. The end result, however, will be better complementary coverage and bibliographic control in all fields. An investigation of bibliographic control in forestry and forest products in the AGRICOLA, CAB ABSTRACTS, and PAPERCHEM databases has been done by Morrow[1].

REFERENCE

1. Morrow, Deanna. "Bibliographic Control in Forestry/Forestry Products: A Comparison of the Scope and Depth of AGRICOLA, CAB/FORESTRY ABSTRACTS and PAPERCHEM." Presented to Special Libraries Association, 69th Annual Conference, Kansas City, June 10-15, 1978 (unpublished).

APPENDIX I. SUBJECT CONTENT OF THE AGRICOLA DATABASE (FORESTRY)

Category Codes:

4500 -- Forestry, (1970-71)
3505 -- Forestry, General
3510 -- Forest Economics and Management
3515 -- Silviculture
3520 -- Forest Industries
4045 -- Plant Physiology and Biochemistry (Forest Trees)
4545 -- Insect Pests and Control (Forest Trees)

APPENDIX II. SUBJECT CONTENT OF THE CAB ABSTRACTS DATABASE (FORESTRY)

Subfiles or File Segments:

FA (FØ) -- Forestry Abstracts
FPA (F1) -- Forest Products Abstracts (1978-)

Oxford Decimal Classification Codes:

11-19 -- Factors of the Environment. Biology.
22-29 -- Silviculture
30-39 -- Harvesting of Wood: Logging and Transport. Forest Engineering
41-49 -- Forest Injuries and Protection
51-59 -- Forest Mensuration. Increment; Development and Structure of Stands. Surveying and Mapping
61-69 -- Forest Management. Business Economics of Forestry. Administration and Organization of Forest Enterprises
71-79 -- Marketing of Forest Products. Economics of Forest Transport and the Wood Industries
81-89 -- Forest Products and Their Utilization
90-99 -- Forests and Forestry from the National Point of View. Social Economics of Forestry

APPENDIX III. SUBJECT CONTENT OF THE FOREST PRODUCTS DATABASE

File Segments and Category Codes:

FUNDAMENTAL WOOD STUDIES
01 -- Biology
02 -- Chemistry
03 -- Physics
04 -- Wood Engineering and Mechanical Properties

MANAGEMENT
05 -- Economic and Financial Management
06 -- Marketing
07 -- Production Management
08 -- Quality Control
09 -- Research Management

PROCESSES
10 -- Drying and Storage
11 -- Energy Committee
12 -- Finishing and Overlaying
13 -- Gluing
14 -- Milling and Machining
15 -- Mechanical Fastening
16 -- Pollution Abatement and Control
17 -- Timber Production

PRODUCTS
18 -- Bark and Residues
19 -- Pallets and Containers
20 -- Construction Plywood
21 -- Decorative Veneers and Plywood
22 -- Furniture Dimensions and Turned Products
23 -- Hardboard and Insulation Board
24 -- Hardwood Lumber
25 -- Laminated Wood, Components and Engineered Structures
26 -- Millwork, Cabinets and Fixtures
27 -- Particleboard and Molded Products
28 -- Pulp and Paper
29 -- Residential and Commercial Buildings
30 -- Softwood Lumber
31 -- Tropical Woods
32 -- Treated Wood Products

APPENDIX IV. SUBJECT CONTENT OF THE PAPERCHEM DATABASE

Subject Areas:

Pulp and Paper Science
Pulp and Paper Technology
Pulp and Paper Raw Materials
Pulp and Paper Products
Packaging
Graphics and Printing
Forestry and Silviculture
Carbohydrate Chemistry
Cellulose Chemistry
Hemicellulose Chemistry
Lignin Chemistry
Wood Extractives Chemistry
Analysis
Engineering
Process Control
Machinery and Equipment
Maintenance
Pollution Control
Water and Power

APPENDIX V. SUBJECT CONTENT OF THE PIRA DATABASE

Subfiles or File Segments:

MMA -- Management and Marketing Abstracts (General)
PBA -- Paper and Board Abstracts
PKA -- Packaging Abstracts
PRA -- Printing Abstracts

Subject Areas:

Adhesives
Binding
Composition
Education and Training
Finishing
Food Packaging
Forecasting
Graphic Arts
Machinery
Occupational Safety
Plastics Packaging
Pollution
Production
Pulps
Recycling
Reprography

Industrial Relations
Inks
Materials

Retailing
Testing

APPENDIX VI. NUMBER OF CITATIONS* IN AGRICOLA RELATING TO FORESTRY, WOOD/FORESTRY PRODUCTS, AND PULP/PAPER PRODUCTS

1 -- /CC 4500 OR 3505 OR 3510 OR 3515 OR 4045 OR 4545 OR 3520	(80,000)
2 -- 3520/CC AND (PULP:/DE OR PAPER/DE)	(3,000)
3 -- 3520/CC NOT 2	(20,700)
4 -- 1 NOT (2 OR 3)	(56,300)

4 -- FORESTRY	(56,300)
3 -- WOOD/FOREST PRODUCTS	(20,700)
2 -- PULP/PAPER PRODUCTS	(3,000)

APPENDIX VII. NUMBER OF CITATIONS* IN CAB ABSTRACTS RELATING TO FORESTRY, WOOD/FOREST PRODUCTS, AND PULP/PAPER PRODUCTS

1 -- /FS FA(FØ) OR FPA(F1)	(50,200)
2 -- /DC (7: OR 8:) AND 1	(18,000)
3 -- 2 AND 861:/DC (Pulps and Paper-making)	(1,000)
4 -- 2 NOT 3	(17,000)
5 -- 1 NOT (3 OR 4)	(32,200)

5 -- FORESTRY	(32,200)
4 -- WOOD/FOREST PRODUCTS	(17,000)
3 -- PULP/PAPER PRODUCTS	(1,000)

APPENDIX VIII. NUMBER OF CITATIONS* IN FOREST PRODUCTS RELATING TO FORESTRY, WOOD/FOREST PRODUCTS, AND PULP/PAPER PRODUCTS

1 -- WOOD OR NOT WOOD (TOTAL FILE)	(12,200)
2 -- FORESTRY/DE,TI	(100)
3 -- 28/CC (Pulp and Paper)	(1,200)

*Rounded to nearest hundred.

4 -- 1 NOT (2 OR 3) (11,900)

4 --	WOOD/FOREST PRODUCTS	(11,900)
3 --	PULP/PAPER PRODUCTS	(1,200)
2 --	FORESTRY	(100)

APPENDIX IX. NUMBER OF CITATIONS* IN PAPERCHEM RELATING TO FORESTRY, WOOD/FOREST PRODUCTS, AND PULP/PAPER PRODUCTS

1 --	PULP OR NOT PULP (Total File)	(134,500)
2 --	FORESTRY/DE,TI	(700)
3 --	/DE FOREST PRODUCTS OR WOOD:	(5,000)
4 --	1 NOT (2 OR 3)	(128,800)

4 --	PULP/PAPER PRODUCTS	(128,800)
3 --	WOOD/FOREST PRODUCTS	(5,000)
2 --	FORESTRY	(700)

APPENDIX X. NUMBER OF CITATIONS* IN PIRA RELATING TO FORESTRY, WOOD/FOREST PRODUCTS, AND PULP/PAPER PRODUCTS

1 --	PBA/FS (Paper and Board Abstracts)	(12,300)
2 --	FORESTRY/DT,TI	(100)
3 --	/DE WOOD: OR LUMBER OR TIMBER	(500)
4 --	1 NOT (2 OR 3)	(11,700)

4 --	PULP/PAPER PRODUCTS	(11,700)
3 --	WOOD/FOREST PRODUCTS	(500)
2 --	FORESTRY	(100)

*Rounded to nearest hundred.

APPENDIX XI. APPROXIMATE NUMBER OF CITATIONS* IN AGRICOLA, CAB ABSTRACTS, FOREST PRODUCTS, PAPERCHEM, AND PIRA RELATING TO THE SUBJECTS OF FORESTRY, WOOD/FOREST PRODUCTS, AND PULP/PAPER PRODUCTS.

	FORESTRY	WOOD/FOREST PRODUCTS	PULP/PAPER INDUSTRY	TOTAL
AGRICOLA	56300	20700	3000	80000
CAB ABSTRACTS	32200	17000	1000	50200
FOREST PRODUCTS	100	10900	1200	12200
PAPERCHEM	700	5000	128800	134500
PIRA	100	500	11700	12300
TOTAL	89310	54200	145600	289200

*Rounded to nearest hundred.

APPENDIX XII. GRAPHIC RESPRESENTATION OF RELATIVE SIZE OF AGRICOLA, CAB ABSTRACTS, FOREST PRODUCTS, PAPERCHEM, AND PIRA IN SUBJECT AREAS OF FORESTRY, WOOD/FOREST PRODUCTS, AND PULP/PAPER PRODUCTS.

SUBJECT / DATABASE	FORESTRY	WOOD/FOREST PRODUCTS	PULP/PAPER INDUSTRY
AGRICOLA	56,300	20,700	3,000
CAB ABSTRACTS	32,200	17,000	1,000
FOREST PRODUCTS	100	10,900	1,200
PAPERCHEM	700	5,000	128,800
PIRA	100	500	11,700

APPENDIX XIII. DOCUMENT TYPES INDEXED BY AGRICOLA, CAB ABSTRACTS, FOREST PRODUCTS, PAPERCHEM, AND PIRA

DOCUMENT TYPE	AGRICOLA	CAB ABSTRACTS	FOREST PRODUCTS	PAPERCHEM	PIRA
Bibliographies	X	X		X	
Books/Monographs	X	X	X	X	X
Conference/Symposia Proceedings	X	X	X	X	X
Government Documents	X		X		
Indexes/Abstracts	X	X			
Patents		X	X	X	
Research/Technical Reports	X	X	X	X	X
Reviews	X	X	X		
Sci/Tech Journals	X	X	X	X	X
Standards/ Specifications		X			X
Theses/Dissertations		X		X	
Trade Journals	X	X	X	X	X
Translations	X	X	X	X	X

APPENDIX XIV. RANGE OF YEARS COVERED BY AGRICOLA, CAB ABSTRACTS, FOREST PRODUCTS, PAPERCHEM, AND PIRA.